Modeling the ultrasonic softening effect for robust copper wire bonding

Andreas Unger, Walter Sextro,
Simon Althoff, Tobias Meyer
Faculty of Mechanical Engineering
University of Paderborn
Pohlweg 47-49, 33098 Paderborn, Germany

Klaus Neumann, René Felix Reinhart
Institute for Cognition and Robotics (CoR-Lab)
Bielefeld University
Universitätsstr. 25, 33615 Bielefeld, Germany

Michael Brökelmann
Hesse GmbH
Vattmannstr. 6, 33100, Paderborn, Germany

Karsten Guth, Daniel Bolowski
Infineon Technologies AG
Max-Planck-Straße 5, 59581, Warstein, Germany

***Abstract*—In power electronics, ultrasonic wire bonding is used to connect the electrical terminals of power modules. To implement a self-optimization technique for ultrasonic wire bonding machines, a model of the process is essential. This model needs to include the so called ultrasonic softening effect. It is a key effect within the wire bonding process primarily enabling the robust interconnection between the wire and a substrate. However, the physical modeling of the ultrasonic softening effect is notoriously difficult because of its highly non-linear character and the absence of a proper measurement method. In a first step, this paper validates the importance of modeling the ultrasonic softening by showing its impact on the wire deformation characteristic experimentally. In a second step, the paper presents a data-driven model of the ultrasonic softening effect which is constructed from data using machine learning techniques. A typical caveat of data-driven modeling is the need for training data that cover the considered domain of process parameters in order to achieve accurate generalization of the trained model to new process configurations. In practice, however, the space of process parameters can only be sampled sparsely. In this paper, a novel technique is applied which enables the integration of prior knowledge about the process into the data-driven modeling process. It turns out that this approach results in accurate generalization of the data-driven model to novel process parameters from sparse data.**

I. Introduction

Ultrasonic wire bonding is an established technology used since decades to connect the electrodes of electrical devices. Because of its flexibility, reliability and cost-efficiency it is widely used for connecting the individual electrodes of micro-electronic chips as well as high power semiconductor modules like insulated-gate bipolar transistors (IGBT). Aluminum wire is preferably used in heavy wire applications because of its robust bonding behavior and low cost.

In the recent years, the growing market of powerful and efficient power modules requires a material with better mechanical and electrical properties. Therefore, copper wire as bonding material is highly desired. The superior material properties of copper compared to aluminium include significantly higher electrical and thermal conductivity, mechanical stability as well as higher interconnection reliability of copper bonds. Therefore smaller chips can be operated at higher temperature but with identical switching power leading to reduced costs and higher yield. For these reasons, a technology change from aluminium to copper is indispensable. Typical application fields of products equipped with copper wire bonds are, for instance, the strongly growing markets of renewable energy and electric vehicles [1].

Copper wire bonding is currently in the state of being established as an alternative interconnection method, mainly in thin wire applications, but recently also in heavy wire bonding of power electronics. Because of the different material properties, the bonding parameters in copper wire bonding differ significantly from those of aluminium wire bonding. Ultrasonic power and the normal bonding forces are about 2 to 3 times higher. The copper wire bonding process also reacts more sensitive to parameter changes. This makes manufacturing of reliable copper bond connections challenging.

In order to increase the reliability of the copper bonds, an adaptation of the bonding parameters at runtime is desired. To this end, self-optimization is employed in the wire bonding machine. Self-optimization gives the machine the ability to adapt its behavior based on the current situation by changing the currently pursued objectives. To do so, optimal compromises between several objectives need to be computed before operation. Therefore, multi-objective optimization techniques are employed [2]. For this purpose, a model of the entire bonding process is required. Although many models of the bonding process have been proposed (see [3], [4]), the interaction of the process parameters is still largely unknown to this day and a sufficiently accurate model is not available. The effect of ultrasound on the wire material is one of those effects that have not yet been modeled successfully. Previous studies assume that the applied ultrasound softens the wire, thus this effect was named *ultrasonic softening*. It has a significant influence on the wire deformation and the bonding process.

In this paper, machine learning is applied to model the ultrasonic softening effect from data. This data-driven modeling extracts regularities of the bonding process from exemplary bonds produced with different process parameters. The trained model can then generalize these regularities to novel process parameters. A typical problem of machine learning techniques is the generalization from only few training examples. Accurate generalization of the data-driven model is achieved in this paper by integrating prior knowledge about the bonding process into the learning. Based on the technique introduced in [5], expert knowledge about the bonding process is rephrased in terms of linear inequalities which then serve as constraints for the learning. This hybrid methodology combines the flexibility of machine learning techniques with the reliability of expert knowledge and physical constraints.

The paper is organized as follows: Section II introduces the physical process and discusses related work. The data-driven modeling technique is introduced in Sec. III. Its application to the ultrasonic softening effect together with the data acquisition are subject of Sec. IV. The paper closes with a brief conclusion in Sec. V.

II. Ultrasonic Softening Effect in Copper Wire Bonding

Ultrasonic wire bonding is a cold friction welding process. The wire is placed under the tip of a slim rod-like bonding tool (see Fig. 1). It is pressed onto the electrode surface with a well-selected normal force causing an initial cold straining at the contact area. A so called ultrasonic transducer generates mechanical vibrations in the ultrasonic range, e.g. 60 kHz, which are transferred by the bonding tool into the welding area. The deformation of the wire and the adhesion between wire and substrate steadily progress during this welding process. Finally a pure inter-metallic compound between wire and electrode is formed at temperatures well below the melting point of the bonding partners. After the first bond the machine forms the so called loop and then the second bond connection is established. After cutting the wire, the tool is lifted and the interconnection process is finished.

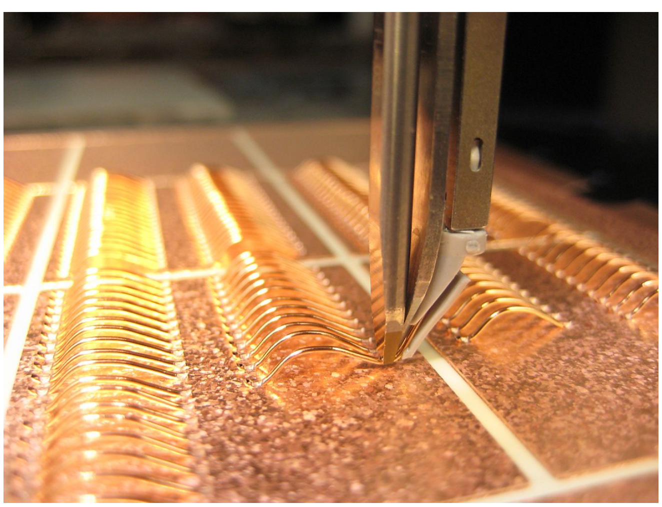

Fig. 1. Copper wires bonded onto a copper substrate.

To build a model of the bonding process it is essential to consider all effects, which are, the static elasto-plastic deformation, the ultrasonic softening effect, and the proceeding adhesion between wire and substrate. This paper focusses on the ultrasonic softening effect. This effect was first presented in [6]. It describes the macroscopic softening of the material under applied ultrasound. The yield strength seems to be lowered so that the material can be deformed at lower mechanical stresses and forces. The ultrasonic softening has a significant influence on the bonding process, because it enables the deformation and growth of the bonding area with reasonable normal forces.

Deformation tests with the bonding machine are conducted to observe the ultrasonic softening effect. The ultrasound is turned on and off during each experiment to examine the effect of the high frequency vibrations. In each experiment, the normal force acting on the wire is increased linearly from 100 cN up to 2800 cN within 1200 ms. The bonding machine and in particular the bonding tool is not perfectly rigid, which makes it necessary to measure its elastic deformation. To do so, the bonding tool is placed on a hard substrate, onto which it exerts the nominal force. The measured elastic deformation of the bonding machine is then subtracted from the total deformation in the regular bonding experiments. Ultrasonic softening and hardening effects are clearly visible in the observed wire deformations shown in Fig. 2. Line 1 depicted in Fig. 2 is a typical force-strain curve for a radially deformed wire without ultrasound. In this experiment, the static normal force only induces a deformation of approximately 13%. In experiments corresponding to line 2, 3 and 4, the ultrasound is applied between 1400 cN and 2100 cN with an altering amplitude from 10 V to 30 V. As can be seen in Fig. 2, the wire softens considerably during application of ultrasound. The amount of deformation increases with higher ultrasound voltages. After shutdown of the ultrasound, the curves show a hardening effect since the gradient of the force-deformation curves are increased compared to line 1 in Fig. 2. As expected, the total deformation after applying ultrasound increases with the ultrasonic amplitude.

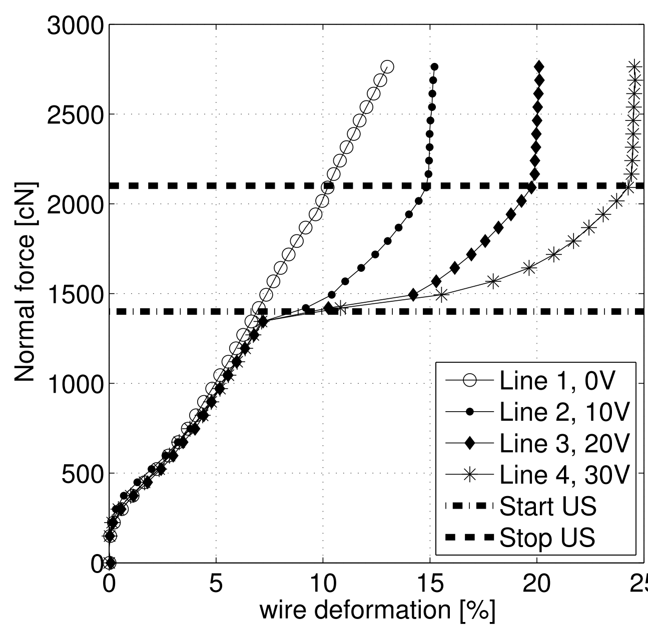

Fig. 2. Deformation experiment showing the ultrasonic softening effect.

Ultrasonic softening is a highly non-linear effect acting on microscopic scale within the metallic lattice. Siddiq et al. [7] modeled this effect in great detail and with high complexity. Within the scope of our research project, a model that can be computed efficiently is required. The research area of machine learning offers techniques that fit flexible models to data and are efficient to compute at the same time. Therefore, such a data-driven modeling approach is favored over a non-linear analytical model in this paper. The aim of the model is to estimate the deformation curves of the wire for different ultrasonic amplitudes and normal force trajectories.

III. Data-Driven Modeling

This section presents a data-driven approach to model the ultrasonic softening effect. First, the model inputs and outputs are introduced. Then, the modeling technique is outlined including the integration of prior knowledge about the bonding process into the learning process.

A. The Copper Wire Bonding Model

A schematic view of the model is shown in Fig. 3. The input of the model is the point in time t of the process, the applied ultrasonic voltage $U_S(t)$ and the normal force $F_N(t)$ between bonding tool and substrate. The output of the model is the wire deformation $D(t)$ at time t. The ultrasonic voltage and normal force are approximately constant during the bonding experiments considered in this paper. Therefore, the model can be understood as a function of time parameterized by the bonding parameters, i.e. ultrasonic voltage and normal force. This leads to a two-dimensional encoding of the bonding process.

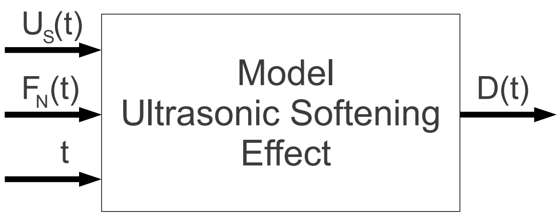

Fig. 3. Schematic model of the ultrasonic softening effect.

The model needs to quantify the impact of different process parameter configurations which are not contained in the data set for learning. Besides this generalization of the training data to novel process configurations, an efficient evaluation of the model for new inputs is an additional requirement. Training can be accomplished offline and is decoupled from the evaluation phase. The following sections introduce a machine learning technique which meets these requirements.

B. Data-Driven Modeling with Extreme Learning Machines

To model the ultrasonic softening effect in a data-driven manner, a so called *extreme learning machine* (ELM, [8]) is applied. The ELM is a feed-forward neural network (see Fig. 4) and comprises three layers of neurons: $\mathbf{x} = (t|U_S|F_N) \in \mathbb{R}^{I=3}$ denotes the input, $\mathbf{h} \in \mathbb{R}^R$ the hidden, and $D \in \mathbb{R}^{O=1}$ the output neurons. The input is connected to the hidden layer through the input matrix $W^{\mathbf{inp}} \in \mathbb{R}^{R \times I}$. The read-out matrix is given by $W^{\mathbf{out}} \in \mathbb{R}^{O \times R}$. For input $\mathbf{x}$, the output is computed by

$$D(\mathbf{x}) = \sum_{j=1}^{R} W_j^{\mathbf{out}} f\Big(\sum_{k=1}^{I} W_{jk}^{\mathbf{inp}} x_k + b_j\Big) \quad , \qquad (1)$$

where b_j is the bias for neuron j, and $f(x) = (1+e^{-x})^{-1}$ the logistic activation function.

The components of the input matrix $W^{\mathbf{inp}}$ and the biases b_j are drawn from a random distribution and remain fixed after initialization. Learning is restricted to the read-out matrix $W^{\mathbf{out}}$. This setup renders learning very efficient because backpropagation of errors like in classical feed-forward networks is not necessary. Although most of the model parameters are randomly initialized and are not trained, it was shown in [9] that the ELM possesses the universal function approximation ability under fairly mild conditions. ELMs feature high efficiency, conceptual simplicity and good generalization capabilities [10].

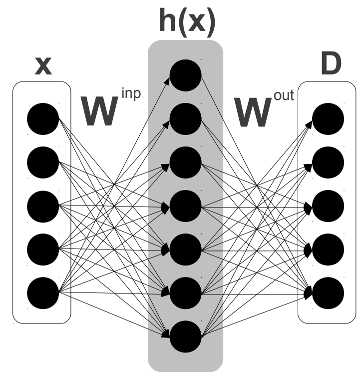

Fig. 4. Extreme Learning Machines (ELMs) are feed-forward neural networks with input, hidden and output layer. Each neuron in the hidden layer computes a weighted sum of the inputs with an additional non-linear activation function. The connection weights from the input to the hidden layer are initialized randomly and remain untrained. Learning is restricted to the read-out weights.

Let $\mathscr{D} = (X,T) = (\mathbf{x}^k, D^k)$ with $k = 1 \dots N_{\text{tr}}$ be the data set for training, where N_{tr} is the number of training samples, X is the collection of input configurations, and T is the matrix of target wire deformations. Supervised learning of the read-out weights $W^{\mathbf{out}} \in \mathbb{R}^{O \times R}$ is accomplished by minimization of the quadratic error functional

$$W^{\mathbf{out}} = \arg\min_{W} \big(\|W \cdot H(X) - T\|^2 + \varepsilon \|W\|^2 \big) \qquad (2)$$

which assumes a Gaussian prior for the learning parameters. The Gaussian prior punishes the growth of the network's read-out weights and is flexibly controlled by the regularization parameter $\varepsilon > 0$. In Eq. 2, $H(X) \in \mathbb{R}^{R \times N_{\text{tr}}}$ is the matrix collecting the hidden layer states obtained for inputs $X \in \mathbb{R}^{I \times N_{\text{tr}}}$, and $T \in \mathbb{R}^{O \times N_{\text{tr}}}$ is the matrix collecting the corresponding target values. The solution to Eq. 2 is given by ridge or Tikhonov regression [11] in a computationally cheap fashion:

$$W^{\mathbf{out}} = T \cdot H(X)^T \cdot \big(H(X) \cdot H(X)^T + \varepsilon \mathbb{I} \big)^{-1} \quad , \qquad (3)$$

where $\mathbb{I} \in \mathbb{R}^{R \times R}$ is the identity matrix. Inputs and outputs are normalized to the range [-1, 1] according to the distribution of the training data. This yields to good learning results without further tuning of parameters.

C. Learning with Prior Knowledge

Learning a well applicable model of the ultrasonic softening effect from few training samples is particularly challenging because only sparse information about the underlying mapping are given. It might be possible that larger parts of the parameter space controlling the bonding process remain uncovered with data. Therefore, there is considerable need for generalization towards regions subject to sparse sampling. Fortunately, prior knowledge about physical properties of the bonding process is available:

(1) The wire deformation is monotonically increasing in time.

(2) The wire deformation is monotonically increasing w.r.t. the ultrasonic voltage because higher voltages lead to a stronger wire deformation.

(3) The wire deformation is monotonically increasing w.r.t. the normal force of the tool. The normal force puts pressure on the wire and thus leads to a stronger wire deformation.

It is possible to rephrase this knowledge in mathematical terms using the model input variable $\mathbf{u} = (t|U_S|F_N) \in \mathbb{R}^3$ to formulate point-wise constraints:

$$\begin{aligned} &(\mathbf{1})\ \ \partial_1 D(\mathbf{u}) = \frac{\partial}{\partial t} D(\mathbf{u}) > 0 : \forall t \in \Omega \ , \\ &(\mathbf{2})\ \ \partial_2 D(\mathbf{u}) = \frac{\partial}{\partial U_S} D(\mathbf{u}) > 0 : \forall U_S \in \Omega \ , \\ &(\mathbf{3})\ \ \partial_3 D(\mathbf{u}) = \frac{\partial}{\partial F_N} D(\mathbf{u}) > 0 : \forall F_N \in \Omega \ , \end{aligned} \tag{4}$$

where Ω is a predefined region in the model's input space.

These constraints are integrated into the learning by taking the partial derivatives of the model's output in a point $\mathbf{u} \in \Omega$:

$$\begin{aligned} \frac{\partial^M D(\mathbf{u})}{\partial u_{m_1} \dots u_{m_M}} &= \sum_j W_j^{\mathbf{out}} \frac{\partial^M h_j(\mathbf{u})}{\partial u_{m_1} \dots u_{m_M}} \\ &= \sum_j W_j^{\mathbf{out}} f^{(M)}(a_j \sum_k W_{jk}^{\mathbf{inp}} u_k + b_j) \cdot a_j^M W_{jm_1}^{\mathbf{inp}} \dots W_{jm_M}^{\mathbf{inp}} \ , \end{aligned} \tag{5}$$

where $m_i = 1,2,3$ denotes the derivatives according to the conditions $(\mathbf{1}),(\mathbf{2}),(\mathbf{3})$. Note that these inequalities are linear in the output parameters $W^{\mathbf{out}}$ irrespective of the form of the constraints for a given point $\mathbf{u} \in \Omega$. This actually defines a quadratic program as first introduced for ELMs in [5].

The read-out weights are then trained by solving this quadratic program optimizing $W^{\mathbf{out}}$ subject to a set of collected point constraints $U_i = \{\mathbf{u}_i^1, \dots, \mathbf{u}_i^{N_\mathrm{u}}\} : \mathbf{u} \in \Omega$ in order to implement the conditions (**1**), (**2**), and (**3**):

$$\begin{aligned} W^{\mathbf{out}} &= \arg\min_W (\|W \cdot H(X) - T\|^2 + \varepsilon \|W\|^2) \\ &\quad \text{subject to: } \partial_i D(U_i) > 0 : i = 1, \dots, 3 \ , \end{aligned} \tag{6}$$

where the matrices $H(X)$ and T again collect the hidden states and targets for inputs X, respectively, and ε is a regularization parameter. Note that the constraints U_i and the input samples X of the training data are not the same. Solving the quadratic program guarantees satisfaction of the given constraints with respect to the discrete inputs $\mathbf{u}$, which is already useful in many applications. It was shown in [5] that a well-chosen sampling of points U_i is sufficient for generalization of the point-wise constraints to a continuous region Ω. The following section introduces a sampling strategy to construct such sets U_i.

D. Sampling Strategy

This section describes a strategy to sample constraints at points $\mathbf{u}$ for the learning by Eqs. 6 which is typically sufficient to generalize the local, discrete set of constraints U_i to a continuous region Ω. The data set D for training and the region Ω where the continuous constraints are supposed to be implemented are assumed to be given a priori.

In a first step ($k = 0$), the network is initialized randomly and trained without any constraints (i.e. the sample matrix $U_i^k = U_i^0 = \emptyset$ is empty). In this case learning can be accomplished by ridge regression – the standard learning scheme for ELMs, see Eq. 3. In the next step, N_C samples $\hat{U} = \{\hat{\mathbf{u}}^1, \hat{\mathbf{u}}^2, \dots, \hat{\mathbf{u}}^{N_\mathrm{C}}\}$ are randomly drawn from a uniform distribution in Ω. Afterwards, the number of samples ν_i that fulfill (**1**), (**2**), or (**3**) are determined according to Eqs. 4. The sampling algorithm stops if more then p percent of these samples fulfill the conditions (**1**)–(**3**), i.e. $\nu_i/N_C > p$. Otherwise, the most violating sample $\hat{\mathbf{u}}$ for each condition is added to the sample pool: $U_i^{k+1} = U_i^k \cup \hat{\mathbf{u}}$. The obtained set of samples is then used for training according to Eq. 6. A pseudo code of the learning procedure is provided in Alg. 1.

Algorithm 1 Sampling Algorithm

Require: data set $\mathscr{D}$, region Ω, counter $k = 0$, three empty sample pools $U_i^k = \emptyset$
Require: $D : \mathbb{R}^I \to \mathbb{R}$ trained with $\mathscr{D}$ by ridge regression
repeat
  draw samples $\hat{U} = \{\hat{\mathbf{u}}^1, \hat{\mathbf{u}}^2, \dots, \hat{\mathbf{u}}^{N_\mathrm{C}}\}$
  $\nu_i =$ no. of samples in $\hat{U}$ fulfilling (**1**)–(**3**)
  if $p > \frac{\nu_i}{N_\mathrm{C}}$ **then** $U_i^{k+1} = U_i^k \cup \arg\max_{\mathbf{u} \in \hat{U}} \partial_i D(\mathbf{u})$
  train ELM with $\mathscr{D}$ and U_i^{k+1} as in Eq. (6)
until $p > \frac{\nu_i}{N_\mathrm{C}} \forall i = 1, \dots, 3$

IV. Experimental Results

The data-driven modeling technique introduced in Sec. III is applied to data acquired from a bonding machine during production of bond connections with copper wire. The generalization ability of ELMs is analyzed systematically.

A. Experimental Setup

A Hesse Mechatronics Bondjet BJ939 bonding machine equipped with a standard wire bondhead is used for data acquisition. This bondhead is designed for bonding with copper wires in the range of $100\,\mu$m to $500\,\mu$m diameter. Tab. I lists the specifications of the bondhead, wire and substrate which are used for all experiments in this paper. In order to obtain

Specification	Value
Bondhead type	RBK01 Back-cut
Transducer Type	60 kHz
Digital Generator Power Output	100 W
Wire Size	$500\,\mu$m (Cu)
Substrate	DCB-thickness: 0.38 mm ceramic, 0.3 mm Cu

TABLE I. Specification of the bondhead, wire and substrate

training data for the data-driven modeling, a grid covering the region of variability of the bonding parameters (voltage $U_S(t)$ and normal force $F_N(t)$) is created. The minimum and maximum ultrasonic voltage as well as the minimum and maximum normal forces are limited by the process. The ultrasonic voltage $U_S(t)$ is varied from 44 V to 52 V and the normal force $F_N(t)$ is varied from 3000 cN to 3800 cN both in 5 equally distant steps. For each grid point, ten individual bond connections are produced. The applied ultrasonic voltage $U_S(t)$ and normal force $F_N(t)$ are approximately constant throughout bonding. The resulting wire deformation trajectories $D(t)$ for each process configuration are recorded, downsampled and averaged. The measured wire deformation is the height reduction during the time interval between touchdown of the tool on the wire and the end of the ultrasonic vibrations.

The experiments are conducted in random order to avoid deviations due to environmental influences. To be able to differentiate between the first and second bond of each interconnection, the type of bond was logged as well. For both of these types, individual models are trained.

B. Learning Setup

ELMs with and without the application of constraints derived from the prior knowledge about the bonding process are trained on the recorded data. To access the generalization ability of the data-driven models, a leave-one-out cross-validation is conducted. That is, training is repeated such that each of the 25 sampled process configurations is once left out from the training set and serves as test scenario. Additionally, the impact of the ELM parameters on the generalization performance is evaluated by changing the hidden layer size $R \in [30, 50, 100]$ and the regularization parameter $\varepsilon \in [10^{-4}, 10^{-6}, 10^{-8}]$. The model performance is evaluated on the training and test set by computing the error between estimated and recorded wire deformation averaged over all time steps. Results are averaged over 10 independent ELM initializations.

C. Results

The cross-validation results for ELM networks equipped with and without prior knowledge are shown in Fig. 5. The best performing network among the networks without prior knowledge has a hidden layer with $R = 50$ neurons and a regularization parameter of $\varepsilon = 10^{-6}$. ELMs with $R = 100$ hidden neurons and a regularization parameter of $\varepsilon = 10^{-8}$ show a typical behavior when learning from few data: They achieve a low training error around $E_{\text{tr}} = 0.002$ but have high corresponding test errors of about $E_{\text{te}} = 0.016$. This big gap between training and test errors indicates strong over-fitting. Applying the learning with constraints, the same networks have a slightly increased training error around $E_{\text{tr}} = 0.003$, but the test error is significantly decreased towards $E_{\text{te}} = 0.0045$. Note that the generalization capability is increased by application of prior knowledge irrespective of the network parameters in this scenario. This demonstrates that the application of suitable constraints alleviates the problem of over-fitting when training data is sparse and that the integration of prior knowledge into the learning facilitates reliable and robust generalization.

Fig. 6 shows the results of the data-driven modeling with and without constraints in more detail. Each cell of the panel corresponds to the process parameters of the bonding machine indicated by the surrounding axes. In each cell, the recorded wire deformation over time for these bonding parameters

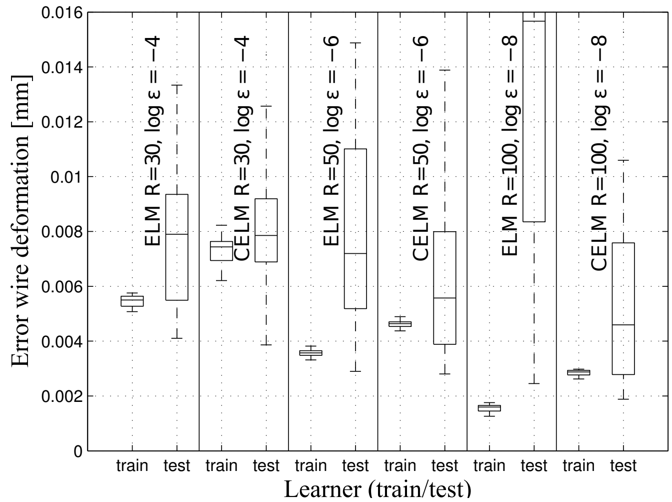

Fig. 5. Cross-validation errors for training and test sets of ELMs without (ELM) and with (CELM) constraints.

is shown by a solid line with standard deviation computed from the ten repeated bonding experiments (gray areas). The generalization performance of models trained with and without prior knowledge are plotted by solid bold and dashed lines, respectively. Due to the high degree of non-linearity, a complex model is required in order to capture the structure of the data. However, a high model complexity, i.e. ELMs with large hidden layers and small regularization constants, are prone to over-fitting and poor generalization when trained on sparse data. In particular, the parameter configurations at the corner and the edge of the matrix are difficult to generalize (see the corners of Fig. 6). This is reasonable since the models need to extrapolate the training examples to these process configurations. Thus, the data-driven modeling without prior knowledge needs a careful fine-tuning of its regularization parameter or network size. On the contrary, Fig. 6 demonstrates that the integration of prior knowledge about the underlying process into the learning by means of linear inequality constraints mitigates the issue of sparse data and results in a significantly increased generalization performance.

V. Conclusion

The ultimate goal of this research is the reliable production of copper wire bond connections under varying conditions. For the in-process adaptation of the bonding parameters, a model-based optimization together with self-optimizing techniques will be applied to achieve this goal in future work. For this purpose, a validated physical model of the process, which can be computed efficiently, is mandatory. Although ultrasonic wire bonding is widely used and the ultrasonic softening effect has been investigated in previous studies, there is still no model of the bonding process or the ultrasonic softening available which fulfills these requirements. This paper proposes a data-driven model of the ultrasonic softening for copper wires for the first time which is accurate and efficient to compute. The issue of poor generalization from sparse data is addressed in this paper by integrating prior knowledge about the ultrasonic softening effect into the learning which then yields to accurate generalization and reduced over-fitting. Supplementing data-driven modeling techniques with prior knowledge about the

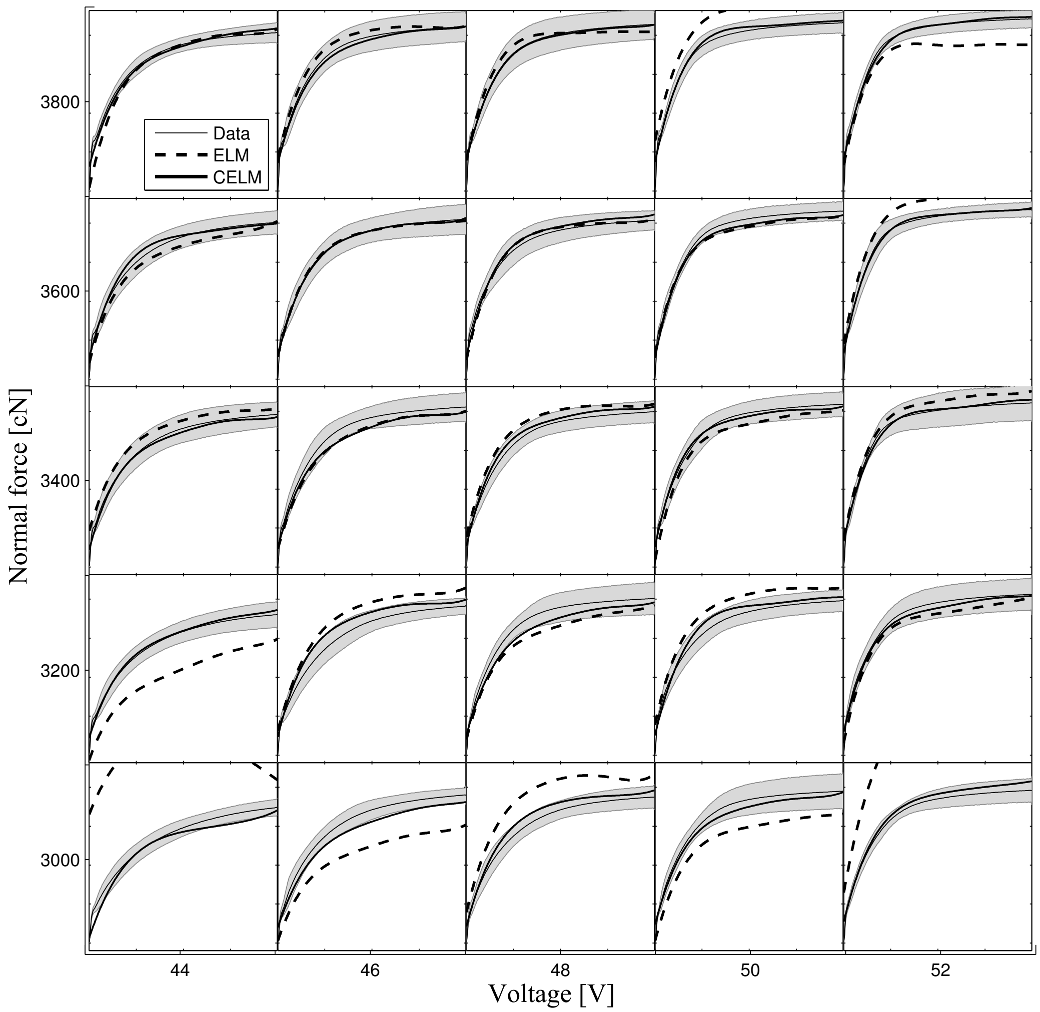

Fig. 6. The cross-validation results of the copper bonding experiment visualized for a ELM model without and with the application of constraints ($R = 100$ and $\varepsilon = 10^{-8}$). The recorded data is depicted in gray and the generalization of the respective parameter configuration is depicted by dashed (ELM) or solid lines (CELM). The constrained learning result in a good generalization performance within the standard deviation of the recorded wire deformations.

underlying process results in highly reliable models and is a promising methodology for a broader range of applications.

ACKNOWLEDGEMENT

This research and development project is funded by the German Federal Ministry of Education and Research (BMBF) within the Leading-Edge Cluster Competition "it's OWL" (intelligent technical systems OstWestfalenLippe) and managed by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the contents of this publication.

REFERENCES

[1] D. Siepe, R. Bayerer, and R. Roth, "The future of wire bonding is? Wire bonding!" *Int. Conf. on Integrated Power Electronics Systems*, 2010.

[2] J. Gausemeier, F. J. Rammig, W. Schäfer, and W. Sextro, Eds., *Dependability of Self-Optimizing Mechatronic Systems*. Berlin, Heidelberg, DE: Springer-Verlag, 2013.

[3] M. Brökelmann, "Entwicklung einer Methodik zur Online-Qualitätsüberwachung des Ultraschall-Drahtbondprozesses mittels integrierter Mikrosensorik," Dissertation, University of Paderborn, 2008.

[4] S. Althoff, J. Neuhaus, T. Hemsel, and W. Sextro, "A friction based approach for modeling wire bonding," *International Symposium on Microelectronics (IMAPS)*, 2013 (unpublished).

[5] K. Neumann, M. Rolf, and J. J. Steil, "Reliable integration of continuous constraints into extreme learning machines," *Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 2013, in Press.

[6] B. Langenecker, C. W. Fountain, and S. R. Colberg, "Effects of ultrasound on deformation characteristics of structural metals," *Defense Technical Information Center*, 1965.

[7] A. Siddiq and E. Ghassemieh, "Thermomechanical analyses of ultrasonic welding process using thermal and acoustic softening effects," *Mechanics of Materials*, pp. 982–1000, 2008.

[8] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in *Proc. Int. Joint Conf. on Neural Networks*, 2004, pp. 985–990.

[9] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.

[10] K. Neumann, C. Emmerich, and J. J. Steil, "Regularization by Intrinsic Plasticity and its Synergies with Recurrence for Random Projection Methods," *Journal of Intelligent Learning Systems and Applications*, vol. 4, no. 3, pp. 230–246, 2012.

[11] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method." *W. H. Winston, Washington, D. C.*, 1977.